Integrative Biology

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/ibiology

Biomass backtracking is a workflow that integrates the cellular context in existing dynamic metabolic models via stoichiometrically exact drain reactions based on a genome scale metabolic model. We provide comprehensive examples for different species and environmental contexts and demonstrate the importance and scope of applications and highlight the improvement compared to common boundary formulations in existing metabolic models.

The presented method allows for the contextualization of dynamic metabolic models based on all available information. We anticipate this to greatly increase their accuracy and predictive power for basic research but also for drug development and industrial applications.

Integrative Biology Accepted Manuscript

Dynamic metabolic models in context: Biomass backtracking

Katja Tummler, Clemens Kühn, Edda Klipp
Theoretische Biophysik, Humboldt-Universität zu Berlin

Abstract

Mathematical modeling has proven to be a powerful tool to understand and predict functional and regulatory properties of metabolic processes. High accuracy dynamic modeling of individual pathways is thereby opposed by simplified but genome scale constraint based approaches. A method that links these two powerful techniques would greatly enhance predictive power but is so far lacking.

We present biomass backtracking, a workflow that integrates the cellular context in existing dynamic metabolic models via stoichiometrically exact drain reactions based on a genome scale metabolic model. With comprehensive examples, for different species and environmental contexts, we show the importance and scope of applications and highlight the improvement compared to common boundary formulations in existing metabolic models.

Our method allows for the contextualization of dynamic metabolic models based on all available information. We anticipate this to greatly increase their accuracy and predictive power for basic research but also for drug development and industrial applications.

Integrative Biology Accepted Manuscript

1 Introduction

Complex metabolic processes play an important role in health and disease. Detailed understanding of their functionality not only helps to unravel cellular economics[1–3] but also stress responses[3], deregulation in diseases such as cancer[4] as well as effective design of microbial factories for biotechnological purposes[5]. Mathematical models in form of ordinary differential equations are powerful tools to understand the intrinsic dynamics and complex functionalities of cellular metabolism[6,7]. Yet, due to limitations in computational power as well as in the availability of high quality data and calibration algorithms, they are usually restricted to a small region of the entire metabolic network or to a simplified description.

Many metabolic pathways have been modeled in great detail, with focus on the central metabolic routes ([8–11] and others) or specific pathways of interest ([12–14] and others). However, the single pathways might show drastically different behavior in changing cellular contexts which the focused models might not be able to account for. On the other hand, a large number of well curated genome scale metabolic models (GSM) have been developed recently[15,16] which are able to predict metabolic fluxes under various external conditions integrating a variety of different experimental data (comprehensive review by Bordbar et al[17]). They, however, lack the potential for dynamic simulations due to the strict steady state assumption. Combining these two powerful approaches will greatly enhance the biological accuracy of the modeled metabolic routes and improve their predictive power.

Methods linking these two types of models have so far focused on improving the predictive power of large scale models. On the one hand, additional constraints derived from a framing dynamic model (dFBA[18]) have been added to GSMs to simulate the transition between different culture phases. Metabolite concentrations can be included to improve the dFBA predictions of metabolic fluxes (MetDFBA[19]). If detailed information is available on the kinetics of single reactions, an approach termed k-OptForce[20] can be used to integrate those to the GSM within a bi-level optimization problem. The approach has been used to significantly improve intervention strategies in industrial strain design. A closely related approach converts all flux-carrying reactions of a whole GSM into a dynamic model with automated and thermodynamically consistent parameter assignment[21]. The resulting in large dynamic models hold the benefit of a broad network coverage giving access to systemic features of metabolism, but their simulation is computationally expensive. A method that tackles the inverse problem of including information of GSMs into detailed and fully kinetic pathway models has been lacking so far.

We present a workflow, *biomass backtracking*, that allows to automatically include the broad knowledge collected in GSMs into smaller detailed dynamic models (DDM) and provide the Matlab scripts necessary for this workflow. Biomass backtracking is based on a simple flux balance analysis (FBA) to calculate all metabolic fluxes occurring at the boundaries of the dynamically modeled network region. It thereby defines a new biomass reaction for the DDM accounting for all systemic requirements of the cell that interact with the modeled pathway under the environmental conditions of interest.

A number of studies used dynamic models of overall cell growth based on nutrient availability to constrain the feasible flux space of a GSM at different time points of batch cultivation [18,22]. The inverse problem of using GSMs to define boundary fluxes of DDMs has not been studied in detail. Of the 169 dynamic ODE models for metabolic pathways in the Biomodels database[23], none has used information on system wide flux distributions, and only 9 take consumption fluxes for biomass production from intermediate species into account. However, GSMs of over 100 species would be

Integrative Biology Accepted Manuscript

available to obtain such information[17].

Defining reactions connecting dynamic metabolic models to cellular context, often in form of biomass reactions, is a complex task that is usually carried out by hand based on own experimental data[9] or crude estimates. In this process it is easy to overlook systemic effects such as side products of distinct synthesis pathways that re-enter the dynamically modeled network part (e.g. pentose phosphate pathway) or the effect of additional nutrients entering at distinct areas of the metabolic network. However, drains into biomass or cellular products and their exact stoichiometry can have dramatic influence on the kinetics of the modeled reactions. If, for example, a dynamically modeled metabolite X is consumed in large quantities for the production of biomass, the reactions consuming it in the dynamic model will have to carry much less flux to maintain the steady state concentration of X than the ones producing it. This strongly influences the kinetic parameters of the associated enzymes as well as all downstream fluxes.

The overall flux distribution has large impact on the systemic behavior of the modeled pathways. Biomass backtracking integrates these effects into one stoichiometrically consistent reaction summarizing all drains from the dynamically modeled network. In addition, it allows transmitting systemic effects of changing conditions such as nutrient restriction or enrichment, modified biomass composition depending on different growth rates, or cell cycle dependent changes in biomass demand directly to the dynamic model and hence allows its simulation in a more realistic cellular context.

Integrative Biology Accepted Manuscript

Backtracking in a nutshell

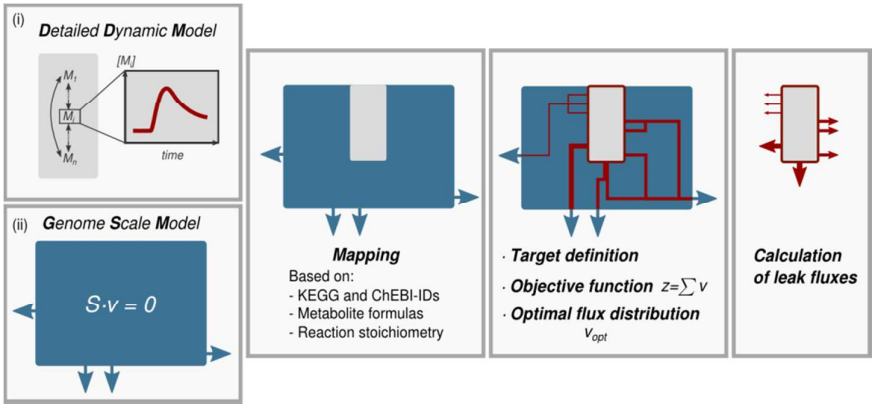


Figure 1: Schematic workflow of the steps carried out during a backtracking run. For detailed description see Online Methods.

The start of backtracking are (i) a DDM covering part of metabolism but lacking a detailed biomass reaction and (ii) a GSM of the same organism (c.f. Figure 1). The backtracking workflow embeds the DDM in the broader context of the GSM and provides a stoichiometrically consistent reaction consuming species of the dynamic model to generate biomass (or any other target metabolite) as defined in the GSM. For this, the overlap between the DDM and the GSM has to be defined (mapping). The mapping will result in a set of reactions and metabolites that is part of both models. We then select a target metabolite from the GSM for whose production we want to quantify the fluxes required to leave the dynamic model. In most cases this will be the biomass, but any other metabolite in the GSM can be used as well, for example if the entity of interest is a specific secreted metabolite. The flux through the reaction consuming the target metabolite is then set to a fixed value, to allow for normalization of the computed leak fluxes. In the next step, an FBA is run to calculate flux distribution in the GSM. At this point, further constraints, such as experimentally measured uptake rates, fluxes or enzyme expression data can be included to constrain the solution space of the GSM. From the obtained flux distribution, we calculate the leak fluxes of each mapped metabolite of the dynamic model as the sum of all fluxes affecting that metabolite (i.e. have non-zero entries in the respective row of the stoichiometric matrix), excluding those that have been mapped to be contained in the DDM.

We provide exemplary implementations to import the models from the commonly used SBML[24] format , map metabolites and reactions shared between both models and define the target metabolite (which will in most cases be the biomass) of the drain reaction. It follows the calculation of leak fluxes v_{leak_i} for each dynamically modeled metabolite M_i of the DDM based on flux distributions in the GSM and their export in form of a lumped leak reaction along with the original model.

Requirements

As described in the main text, biomass backtracking equips a DDM covering a part of the metabolism with a detailed biomass reaction based on a GSM of the same organism. The stoichiometrically consistent reaction consumes species of the dynamic model to generate biomass or any other target metabolite as it is defined in the GSM.

Integrative Biology Accepted Manuscript

The DDM and GSM are imported from a commonly used format (SBML or a COBRA Matlab model directly). The metabolites and reactions of the DDM are identified in the GSM (mapping) based on provided annotations. This assembly is followed by the calculation of leak fluxes v_{leak_i} for each dynamically modelled metabolite of the DDM based on the distribution of fluxes in the GSM, which are calculated by the overall minimization of fluxes. From those leak fluxes a biomass or drain reaction is compiled and can be exported alongside the DDM.

Model Import

In our workflow, we use the COBRA toolbox[25] which is designed for the analysis of genome scale metabolic models. The toolbox supports the import of SBML models via the SBML toolbox for Matlab ([26], sbml.org/Software/SBMLToolbox), which was applied to a number of published genome scale metabolic reconstructions. When importing metabolic reconstructions from SBML, it is crucial that the annotations are imported correctly because they are used to compute the overlap with the DDM. In case the annotations were not correctly transferred from the SBML file to the COBRA model structure, we provide a function to refill the annotations and metabolite formulas via a database query where necessary (function bktr_refillChebiAndKEggIDs.m). The DDMs of interest were imported via the same functions. The generated COBRA-format version of the DDMs cannot be used to simulate the model's transient behavior, but includes all topological information as well as the provided annotations, which is sufficient information for the backtracking workflow. The kinetic equations are stored in the SBML toolbox version of the DDM and are reinserted during the export from Matlab to SBML after the backtracking.

Mapping

The next step is to map which part of the GSM is represented in the DDM. We provide a mapping function bktr_mapping.m to match the annotations of the dynamic and genome scale model and summarize the found mapping in a Matlab structure. The mapping of metabolites is based on ChEBI [27] and KEGG [28] identifiers as well as the metabolite formulas. If more than one matching annotation is found, a manual selection is required. For metabolites of the DDM that are the sum of several metabolites (*lumped metabolites*) a list of matches can be defined. Reactions are mapped based on KEGG identifiers, EC-numbers, associated genes and their stoichiometry. The stoichiometry mapping builds on the preceding mapping of the metabolites and finds reactions in the GSM that have the same substrates and products as the reactions in the dynamic model. The stoichiometry does not need to match exactly to catch also dynamic reactions that neglect co-factors such as hydrogen atoms or phosphate. Again, a selection from a list of found matches is possible and a list of corresponding reactions can be defined for lumped reactions. It is also possible to exclude species and reactions of the DDM from the calculations, which becomes necessary if the DDM already includes dummy reactions for storage compounds or ATP consumption. These reactions and metabolites can be tagged with an ‘rm’ -flag in the mapping process and will be ignored for the calculations.

For the mapping of both metabolites and reactions special attention should be paid to the compartments in which they are found. Many GSMs include several cellular compartments whereas DDM typically cover only a subset of them and are often not well annotated. During the mapping the correct entity of the GSM should be selected from the list of found matches. Lumping of ambiguous entities as for example ATP over several compartments should be handled with care, as it will neglect effects of concentration differences between the compartments which will directly impact the kinetics of the resulting DDM.

Integrative Biology Accepted Manuscript

The effort required to map the DDM on the GSM heavily depends on annotation density of the two models. For incompletely transferred annotations during the SBML import of the GSM model, we provide the function bktr_RefillChebiAndKeggIds to add missing annotations and metabolite formulas from the SBML-toolbox version of the model and the ChEBI and KEGG homepages directly.

Targets

The target of the backtracking can be any reaction or metabolite of special interest contained in the GSM, in most cases this will be the 'biomass' metabolite or the 'growth' reaction. In the function bktr_target, the flux through the selected reaction or a newly generated reaction consuming the selected target metabolite will be fixed in the GSM to be used in the following steps. The fixing allows calculating ratios of the consumption fluxes for each species of the dynamic model needed to produce the set amount of target metabolite.

Objective Function and Flux Optimization

As we fix the flux through the consuming reaction of the target metabolite, the objective of the flux balance analysis (FBA) to calculate the distribution of cellular fluxes is taken to be the sum over all fluxes. The choice of the appropriate objective function is a key point for the predictive power of constraint based metabolic modeling. In many cases, maximizing the growth rate is used as biological and evolutionary objective of the cell, originating from the beginnings of FBA of constraint based metabolic models [17,29]. As we are interested in how many precursor molecules are needed to generate a fixed amount of biomass (or target metabolite), we use the more general overall flux minimization corresponding to a minimization of cellular enzyme synthesis costs, which has been shown to be in good accordance with in-vivo fluxes, at least in bacteria[30,31]. Depending on the organism and the experimental conditions, this objective might need to be adapted.

The function bktr_prep_lump.m generates an irreversible version of the GSM with the minimum flux objective based on the previously generated model with fixed target flux. This function can potentially be amended for the use of different objective functions.

Leak calculation

For the GSM fulfilling the steady state assumption $S \cdot v = 0$, with the stoichiometric matrix S we define v_i as the flux of all reactions that have been mapped between the GSM and the DDM and v_j as all fluxes only present in the GSM (with $v = [v_i\ v_j]$). We define *leak fluxes* v_{leak_i} as the sum of fluxes v_j of all reactions that consume or produce a metabolite M_i of the DDM and are themselves not part of it multiplied with their respective stoichiometric coefficient $\vartheta_{i,j}$.

$$v_{leak_i} = \sum_{j=1}^{n} \vartheta_{i,j} \cdot v_j$$

Their calculation from the found optimal flux distribution is based on the mapping of the DDM on the GSM. For reactions and metabolites with a one-to-one mapping the leak fluxes can be easily calculated as a simple weighted sum. However, the level of simplification in the dynamic model gives rise to several special cases:

Leak fluxes for *lumped metabolites* are calculated as the sum over all leak fluxes of all mapped species of the metabolite pool. Reactions interconverting between the pool species will cancel out in the summation, but their contribution to the turnover of cofactors will still be represented in the optimal flux distribution. *Lumped reactions*, for example one single reaction representing a linear pathway, have to be treated more carefully. As they also bridge several intermediate metabolites, the leaks of those will not be represented in the DDM. However, those could be essential sources for

Integrative Biology Accepted Manuscript

biomass precursors and neglecting their leaks could introduce large errors in the flux distribution of the DDM. We calculate a summed stoichiometry of the lumped reaction from the optimal flux distribution with amended stoichiometric coefficients of substrates and products. For each lumped reaction, a '*waste*' term is generated including all intermediate leaks from the bridged metabolites. The function bktr_run.m can be used to perform the FBA and calculate the leak fluxes.

Variability analysis (optional)

The calculated leak fluxes are the result of a simple FBA and hence only represent one possible flux distribution that is consistent with the set constraints. To assess the variability in the calculated leak fluxes a standard flux variability analysis (FVA[32]) analysis can be used (function bktr_leakFVA.m). The variability in the leak fluxes is the calculated as the lower and upper bounds lb_{leak} and ub_{leak}

$$lb_{leak_i} = \sum ub_{cons_i} + \sum lb_{prod_i}$$

$$ub_{leak_i} = \sum lb_{cons_i} + \sum ub_{prod_i}$$

where ub_{cons_i} is the upper flux variability bound of a reaction that consumes the dynamically modeled metabolite M_i, i.e. that has a negative entry in the ith row of the stoichiometric matrix of the GSM. lb_{leak_i} is be the lowest possible value (lower flux bound) for the leak flux from M_i.Since the leak fluxes are a sum of several fluxes, the variability can add up and give rather large bounds for the leak stoichiometries. FVA has the drawback that only the individual variability for each reaction is calculated and the interdependency of the feasible flux bands is lost.

Reasons for variability can either be the insufficient quality of the GSM, the nature of the objective function (e.g. the overall sum of fluxes can be achieved by several combinations of fluxes) or actual variability in the given network. The robustness of the found leak flux distribution may also be assessed by allowing small deviations from the found optimum value of the objective function.

Model Export

The dynamic model with amended stoichiometries for lumped reactions and the newly generated biomass reaction can be exported to SBML via the SBML toolbox and the provided function bktr_export.m.

Integrative Biology Accepted Manuscript

3 Results

In the following we present various showcases to demonstrate the capabilities and limitations of biomass backtracking. We describe the general workflow using yeast glycolysis as an example and highlight additional insights that were obtained by biomass backtracking. Thereafter we present different application examples focusing on the inclusion of experimental data as well as species specificity.

Showcase 1: Embedding a glycolytic model into the GSM context in yeast

We combined the DDM of yeast glycolysis by Hynne et al.[8] (available from Biomodels as BIOMD0000000061, c.f. Figure 1A) with the current version of the yeast consensus metabolic network (Yeast7[33], available via yeast.sourceforge.net). The Hynne model was originally intended to study glycolytic oscillations in normal and respiration inhibited yeast cells and includes the reactions of glycolysis down to pyruvate production from which acetaldehyde and ethanol are produced. Some reactions are lumped. We removed the reactions concerning the impact of cyanide from the model, since we are focusing here on the unperturbed state. Biomass production is represented by a drain reaction into storage compounds from glucose-6-phosphate. In addition, the model can excrete glycerol, acetaldehyde and ethanol to the extracellular space. Other intermediate metabolites are not consumed for the generation of biomass.

Leaks to produce the complete biomass

We first ran a leak calculation with the complete biomass as target ('s_0450' in Yeast7, Figure 2A). Since the Hynne model is well annotated, it could be easily mapped using our workflow (for details see Supplementary Text S4). As in the original DDM, the leak fluxes for biomass production mainly originate from G6P but a smaller amount also drains from fructose-6-phosphate towards the pentose phosphate pathway and the production of nucleotides and amino acids. From the pentose phosphate pathway, part of the flux is re-entering the DDM, such that dihydroxyacetone-phosphate and glyceraldehyde-3-phosphate have a positive leak flux. That in turn means that the downstream reactions will have to carry a higher flux than the ones upstream of the two components, which is not accounted for in the original model.

The reaction lpPEP is a lumped reaction representing the conversion of bisphosphoglycerate to phosphoenolpyruvate with net-production of one ATP. Lumping these reactions simplifies the DDM but skips the intermediate metabolites 2- and 3-phosphoglycerate. The backtracking workflow identified that the lumping neglects a leak from 3-phosphoglycerate and proposes an amended stoichiometry for the reaction, where the hydrolysis of one mole bisphosphoglycerate produces only 0.93mol phosphoenolpyruvate. The skipped leak from 3-phosphoglycerate is stored in the 'waste'-term provided by the workflow for each lumped reaction. For more detailed discussion of the issues of changing model topologies, e.g. by lumping reactions see Supplementary Text 5. Finally, the largest leak flux is draining from PEP to be shuttled into mitochondria to produce ATP via the respiratory chain and other biomass precursors.

Splitting in single biomass compounds - towards modularization.

To be able to adjust for changing biomass compositions under different environmental conditions, we exemplarily split the biomass into its main constituents based on experimental measurements[34] (Figure 2B, Supplementary text S4). We assessed how much of each dynamically modeled precursor from the Hynne model was required to synthesize each biomass component. Most of the acquired carbon drains from fructose-6-phosphate and pyruvate but many other intermediates contribute to

Integrative Biology Accepted Manuscript

building of the biomass components. Pyruvate is required for the synthesis of all compounds since it provides the additional energy needed to synthesize biomass components via the TCA and oxidative phosphorylation (c.f. Figure 2D and below). In addition, it is shuttled to the TCA cycle from which especially amino acid precursors are withdrawn.

Additional Features

Using the GSM allows for the quantification of additional fluxes that can be useful in the modeling process as well as for consistency checking. In the backtracking of individual biomass components, this permits quantifying how much the production of each compound contributes to the required uptake of nutrients and cofactors (Figure 2C). In a DDM, the consumption or production of additional metabolites could be included in the newly defined biomass reaction to correctly balance the mass flow, for example in the case of CO_2 release, or to model the dependencies on nutrient or oxygen uptake with correct stoichiometries. Likewise, the effect of restrictions in the uptake rates, such as reduced availability of nutrients or oxygen, on different production ways can be included in the DDM.

ATP and cofactor turnover

Backtracking also provides leak fluxes for cofactors and energy compounds, such as ATP and NAD(P)H. For these highly connected species, turnover rates are difficult to quantify. However, they have a system wide impact on the DDMs, contributing to many kinetic reaction rates. Finding consistent turnover rates for those compounds is hence essential to ease dynamic model calibration and ensure biological plausibility. Balancing of cofactor turnover is exacerbated if different compartments have to be taken into account. Turnover rates and concentrations of the same metabolites might differ significantly between compartments but can in many cases only be measured in bulk for the whole cell. Biomass backtracking can use the information from the GSM to calculate the part of turnover taking place in the compartment(s) contained in the DDM and hence enable a more precise balancing.

We calculated the ATP leaks for the Hynne DDM for the production of individual biomass constituents. Apparently, ATP investment was needed for several biomass precursors. Glucan production, for example, mostly drains glucose-6-phosphate requiring ATP in upper glycolysis (c.f. Figure 2D red bars representing $-v_{leak}$). Hence, for this target compound, we saw an additional leak of pyruvate towards the TCA and respiratory ATP production. Because the Yeast7 model possesses compartments, the backtracking only takes into account a fraction of the overall ATP turnover, i.e. all reactions changing the concentration of cytoplasmic ATP which is part of Hynne's DDM. From the cytoplasmic perspective, we only see pyruvate leaving the compartment and ATP entering from the mitochondrial compartment after oxidative phosphorylation. If the respiratory ATP production is larger than the glycolytic ATP investment, we observe a positive ATP leak. This is the case for most biomass precursors originating from upper glycolysis. For biomass precursors that are mainly derived from downstream intermediates, such as many amino acids, the glycolytic ATP production can contribute to the synthesis costs.

For comparison, the overall ATP turnover required for synthesis of the biomass component is shown in Figure 2D (gray bars). The difference between this and the glycolytic leak is the contribution of the respiratory consumption of the drained pyruvate (Figure 2D light bars).

Integrative Biology Accepted Manuscript

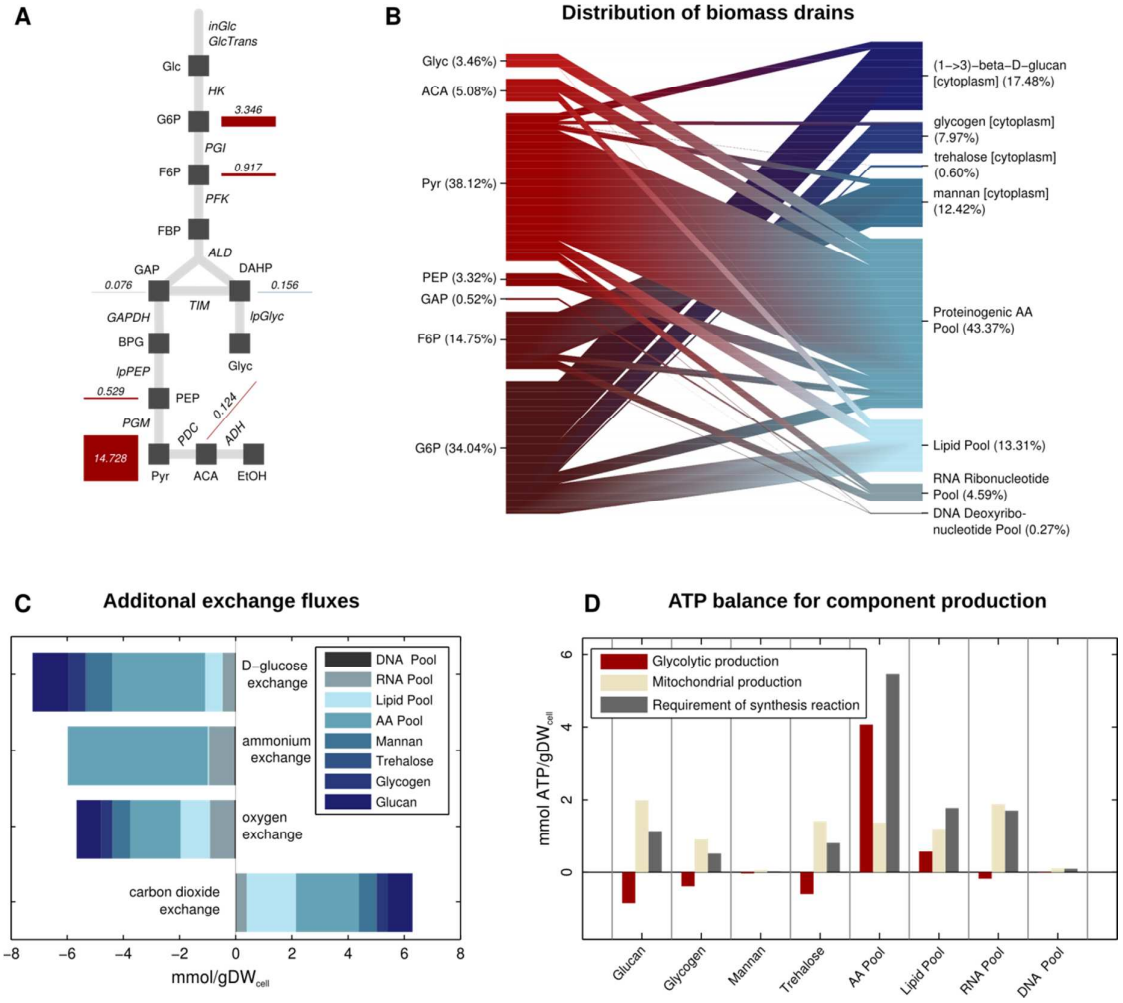


Figure 2: **A** Topology of the original Hynne [8] model and calculated leak fluxes (mol/mol biomass, red: consumed from the dynamic model, blue: released into the dynamic model), **B** How much of the carbon goes where? Percentage of carbon used for each biomass component weighted with biomass molarities of [34] and the number of carbon atoms in each precursor, **C** Additional exchange fluxes required for the production of the biomass components **D** Balance of ATP investment in the different biomass components, red bard representing the glycolytic production of ATP equal to $-v_{leak,ATP}$ of the Hynne model.

Abbreviations: G6P - glucose-6-phosphate, F6P - fructose-6-phosphate, DHAP - dihydroxyacetone-phosphate, GAP - glyceraldehyde-3-phosphate, BPG - bisphosphoglycerate, PEP - phosphoenolpyruvate, ACA - acetaldehyde, EtOH - ethanol, glyc - glycerol

Integrative Biology Accepted Manuscript

Comparing fluxes and dynamics

For the Hynne model we found that the simple boundary definition has neglected a number of leak fluxes from the intermediate metabolites that will have a significant effect on reaction kinetics. To strengthen these findings we calculated leak fluxes for 3 further high quality DDMs using backtracking and compared them to the boundary fluxes present in the models (Figure 3A and Supplementary Text S5). We found that for all models backtracking predicts leak fluxes that were neglected in the DDMs. The closest match to the backtracking leaks was found for the model of Chassagnole et al.[9], whose boundary fluxes are based on detailed experimental data. However, the vast majority of currently available metabolic DDMs has less detailed boundary fluxes and could potentially be improved by more accurate leak fluxes from backtracking (Figure 3B).

To exemplarily asses if a detailed biomass reaction based on biomass backtracking can improve the predictive power of a dynamic model, it was added to the model of van Eunen[35] et al.. We chose this very high quality model as it provides dynamic datasets, is well documented and reproducible. All dummy consumption reactions were removed and the new biomass reaction from the backtracking was added (for methodological details see Supplementary Text S6 and tables therein). A crucial part of the addition of the new biomass reaction is thereby to define its kinetic law, which needs to avoid the biomass precursors to attain negative values while being numerically feasible. In any case, this kinetics will be a stark simplification of the involved biological processes. We utilized simple scaled formulations that avoid unnecessary mechanistic complexity. We find that, with different fitting algorithms, the addition of the biomass reaction can improve the already very good model of van Eunen et al. by 5 – 15% in terms of the overall deviation of the model trajectories from the experimental data compared to a naïve refit of the original model (for detailed analysis see Supplementary Text S6).

Integrative Biology Accepted Manuscript

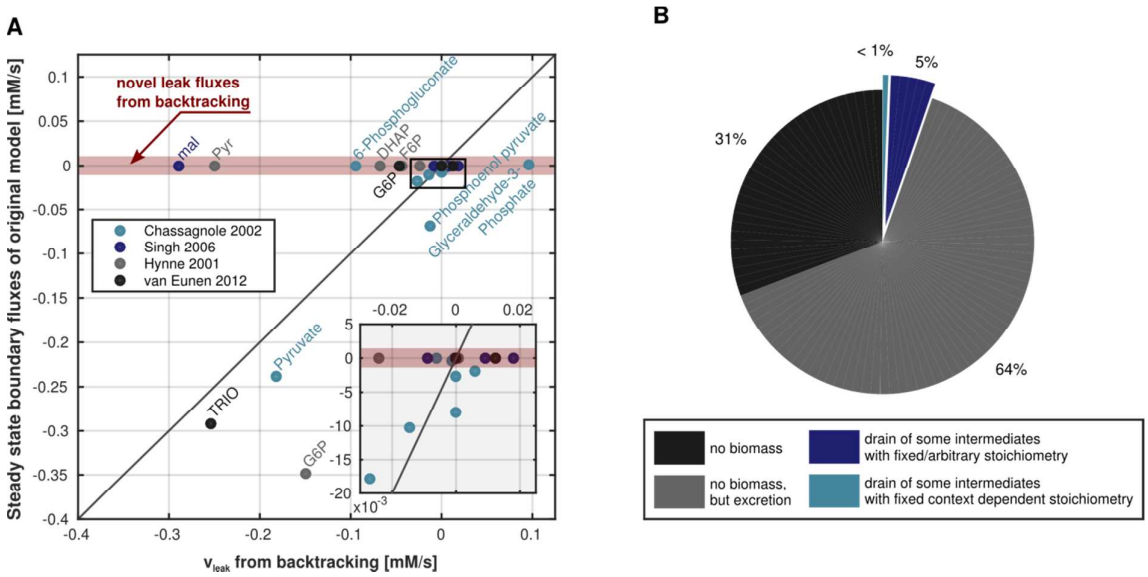


Figure 3: Comparison of leak fluxes of existing dynamic models to the results of backtracking. **A** Leak fluxes were calculated for two dynamic model of *E.coli* from Chassagnole et al.[9] (BIOMD0000000051) and Singh et al.[12] (BIOMD0000000221) based on the iJO1366[36] GSM, and the dynamic models of *S. cerevisiae* from Hynne et al.[8] (as above, BIOMD0000000061) and van Eunen et al.[35] (MODEL1403250001) based on Yeast 7[33] and compared to the steady state fluxes that leave the respective models as simulated in CoPaSi[37]. Fluxes in the red area were predicted by backtracking but absent in the dynamic models. Fluxes outside the depicted flux space can be found in Supplementary Table 2. **B** Summary of the models of all dynamic metabolic models in the Biomodels database with regard to their definition of boundary fluxes. For more detailed description see also Supplementary Text S5 and Supplementary Table2.

Integrative Biology Accepted Manuscript

Showcase 2: Oxygen consumption in yeast

As an example how changing environmental conditions can influence the leak fluxes of a dynamic model, we investigated leaks from a hypothetical DDM of yeast glycolysis under different oxygen availabilities based on chemostat data [38]. The experimentally measured uptake and secretion rates of oxygen, glycerol, ethanol and CO_2 were used to confine the Yeast7 model according to the five tested oxygen availabilities (Figure 4, inset). The glucose uptake rate was left variable and found to be in good accordance with the measured rates in the simulation results (Figure 4, inset, 2nd column). The altered physiology of the limited oxygen availability could be transmitted to the leak fluxes of the glycolytic intermediates (Figure 4). With decreasing oxygen availability, more glucose needs to be taken up to sustain the chemostat growth rate of 0.1 h^{-1}, and hence also the glycolytic reactions will have to carry a significantly higher flux, which the dynamic model needs to account for. A DDM of this process would have to be equipped with appropriate kinetics to reproduce all of the observed leak fluxes, potentially also accounting for changing levels of metabolic enzymes. The information on the necessary leak fluxes strongly reduces the feasible parameter space for model calibration.

Integrative Biology Accepted Manuscript

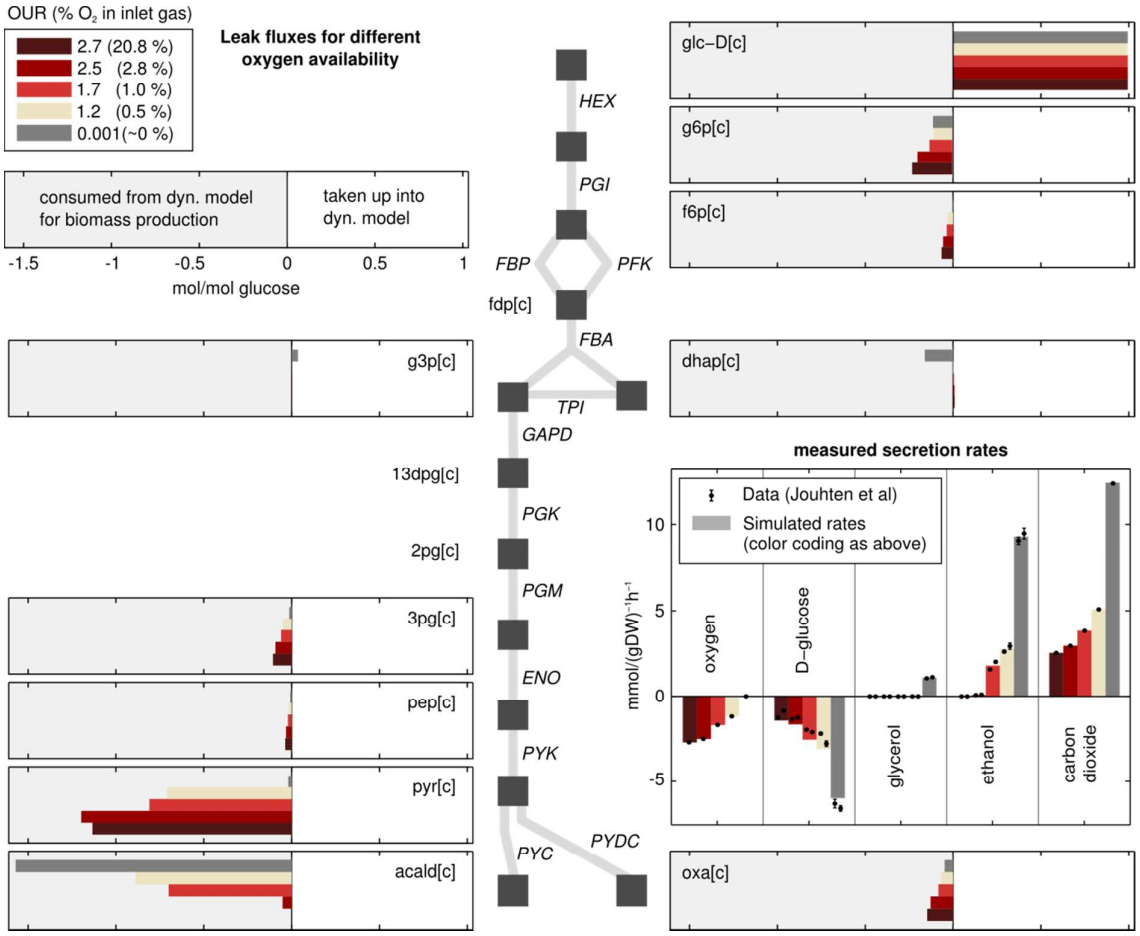


Figure 4: Leak fluxes from glycolysis for different oxygen availabilities in a chemostat experiment. Inset shows the measured uptake rates of Jouhten et al ([38], mean ± SEM) along with the resulting implemented constraints on the Yeast7 model. For details see Supplementary Text S3 (Ex3). As we focus on the redistribution of the leak fluxes with decreasing oxygen availability the fluxes were normalized to the glucose influx (cmp. inset and leak fluxes of intracellular glucose) for comparability. The Yeast7 model cannot adequately represent anaerobic growth[22], the anaerobic case is simulated by a very low oxygen uptake rate (OUR) that still gives a feasible solution in the FBA. Changes in the biomass composition due to the changing oxygen availability could not be accounted for due to the lack of respective data.

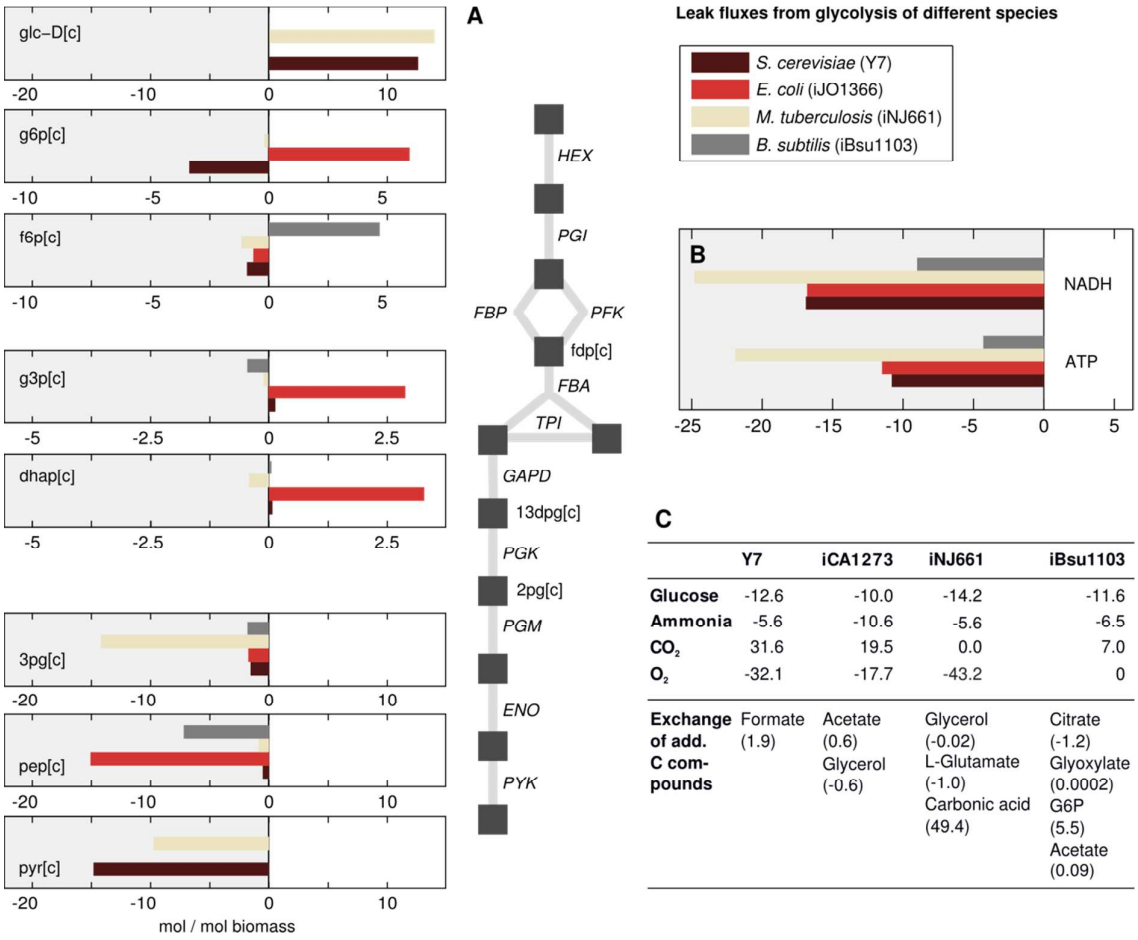
Integrative Biology Accepted Manuscript

Showcase 3: Drain from glycolysis in different organisms

Not only changing environmental conditions can have an impact on the leak fluxes, but also organism specific topological and functional differences of the metabolic network. As an example, we compare the leaks from a hypothetical glycolysis DDM in the species *Saccharomyces cerevisiae*, *Escherichia coli*, *Mycobacterium tuberculosis* and *Bacillus subtilis* (Figure 5, Supplementary Table 1). For all four organisms one or more well curated GSMs are available. For this investigation the Yeast7 as above[33]), iCA1273[39], iNJ661[40] and iBsu1103[41] GSMs were used. The media composition was chosen to represent standard minimal cultivation media for each species with glucose as sole carbon source (Supplementary Table 1). All models were imported, annotated, and mapped as described in Supplemental Material. Since glycolysis is a highly conserved pathway all reactions could be mapped in all organisms (mapping summarized in Supplementary Table 1). The calculated leak fluxes proved to be significantly different between the 4 species (Figure 5A).

In yeast and *M. tuberculosis*, glucose is taken up directly such that the glucose node has positive leak flux. *E.coli* and *B. subtilis*, instead, acquire glucose via phosphotransferase systems (PTS), which use PEP hydrolysis to pyruvate to directly phosphorylate glucose to glucose-6-phosphate during uptake. Hence, the influx into glycolysis is only seen at the level of glucose-6-phosphate for *E.coli*. *B. subtilis,* in addition, processes incoming glucose via β-glucose-phosphates which are isomerized to fructose-6-phosphate, such that the flux enters the glycolytic model at the fructose-6-phosphate node. The PTS are also the reason why the main leak flux from lower glycolysis is draining from PEP for *E.coli* and *B. subtilis*: it is directly hydrolyzed in the PTS reaction; the pyruvatekinase reaction does not carry flux. For yeast and M. tuberculosis the main flux drains from pyruvate to feed the TCA cycle and respiratory ATP production. Each organism also addresses different energetic requirements to its glycolytic pathway (Figure 5B).

Integrative Biology Accepted Manuscript



C

	Y7	iCA1273	iNJ661	iBsu1103
Glucose	-12.6	-10.0	-14.2	-11.6
Ammonia	-5.6	-10.6	-5.6	-6.5
CO_2	31.6	19.5	0.0	7.0
O_2	-32.1	-17.7	-43.2	0
Exchange of add. C com-pounds	Formate (1.9)	Acetate (0.6) Glycerol (-0.6)	Glycerol (-0.02) L-Glutamate (-1.0) Carbonic acid (49.4)	Citrate (-1.2) Glyoxylate (0.0002) G6P (5.5) Acetate (0.09)

Figure 5: Comparison of the glycolytic requirements in different organisms. **A** Leak fluxes of the single components, **B** NADH and ATP requirements to produce one mole of biomass for each organism, **C** Comparison of additional uptakes and secretion rates.

Integrative Biology Accepted Manuscript

Showcase 4: Using backtracking to correct model topologies

Biomass backtracking is especially useful in the process of model development and in integrative modeling approaches that try to link several aspects of the cellular physiology. In those processes it is essential to iteratively improve DDMs based on the available data and new biological insights which will in many cases include a change of the model structure. On the other hand, the insight gained from the previous DDM versions should be usable for the improved version, for which consistent leak fluxes play a major role. The backtracking workflow provides the possibility to calculate leak fluxes that are consistent with the previous DDM versions as well as with the overall biological context.

Conceptually, three different scenarios can occur during the iterative development of DDMs: DDMs can be *split* into smaller individual submodels, *simplified* or *extended*. We demonstrate each case exemplarily using yeast glycolysis. We exemplify all three cases using the dummy DDM of glycolysis from above. For comparison we show the leak fluxes for a full glycolytic model in Figure 6A. The flux entering the system at the glucose node is consumed from G6P, F6P, 3PG, PEP and PYR and small fluxes are reentering glycolysis at the level of the triose phosphates.

Splitting models

Splitting DDMs holds several benefits. Smaller models are usually easier to handle and to overview for the developer. The simulation time and thus also the time for parameter estimation routines increases with the size of the model. In addition, numeric problems can be omitted by splitting models in different time scales.

We split our hypothetical DDM in upper and lower glycolysis and consistently recalculated the leaks for each submodel (Figure 6B). With G3P dehydrogenase (GAPDH), we remove the only NADH dependent reaction and hence obtain a system that is independent of the highly linked co-factor – and hence much easier to simulate and parameterize. For each obtained sumodel a separate ATP balance can be calculated: Upper glycolysis consumes ATP with a leak flux of 20.98 mol/mol biomass, i.e. ATP has to be invested in this part of the system. Lower glycolysis produces ATP via the *PGM* and *PYK* reactions, resulting in a leak flux of -31.76 mol/mol biomass. The submodels could now be calibrated individually with much less computational effort while keeping the overall mass and energy balance of the system constant.

Simplifying models

Wherever the data coverage is not sufficient enough to allow the estimation of kinetic parameters with sufficiently high confidence, DDMs should be simplified to reduce the number of degrees of freedom (i.e. parameters)[42]. The reduction of DDMs can also be based on time scale separation (lumping of fast reactions [43]), topological features of the model[44]. Simplified models hold the same benefits as smaller models in terms of computational effort and clarity.

The topology of the DDM can be adapted to the available data by lumping reactions and/or metabolites. In our example, we generated a DDM with three lumped reactions and one lumped metabolite pool (Figure 6C). Merging reactions is more complex if the resulting lumped reaction skips intermediates that contribute to the biomass production. In the above example, it would be wrong to assume that all acquired glucose molecules are converted to fructose-1-6-bisphosphate (FBP) by the first lumped reaction *LR1*, as part of them is also used elsewhere (e.g. the pentose phosphate pathway and the direct production of biomass). The subsequent *FBA* reaction would hence have to carry much less flux, which directly impacts on the estimation of its kinetic parameters. The backtracking workflow resolves this inaccuracy by introducing 'waste terms' for each lumped

Integrative Biology Accepted Manuscript

reaction in the DDM as well as an amended stoichiometry. Importantly, the stoichiometry of the lumped reactions can be different between two distinct biological scenarios. In our example, *LR1* is consuming one glucose molecule but produces only 0.66 molecules FBP and a waste term as shown in the table in Figure 6C. The occurrence of large waste terms for lumped DDM reactions may also help in the model design process as it hints to potentially important, but not yet included system features.

For pooled species, the leak fluxes of all mapped metabolites belonging to the pool will be added up. The reactions interconverting between the pooled species can be taken into account, if they have been mapped, which is especially important if they also turnover co-factors and ATP.

Extending models

During model development, it is often necessary to extend DDMs with additional reactions or whole pathways, for example if new data is generated or additional reactions become interesting under changing external conditions. To be able to utilize the calibration of the previous DDM version it is essential to obtain consistent boundary fluxes for the extension that can redistribute the original leak flux from the extended part of the system, but conserve the balance for the unchanged part of the DDM. Especially if highly connected metabolites are included in the network extension, the calculation of consistent leak fluxes can become complicated and unintuitive.

In our example, we extend upper glycolysis with the pentose phosphate (Figure 6D). The leak fluxes for G6P, F6P and G3P need to be changed and the resulting difference in the leak fluxes is consistently distributed over the added metabolites of the pentose phosphate pathway. The backtracking also yields information about the NADPH requirement from the extended model (-1.018 mol NADPH/mol biomass).

In our example we use an intensely studied pathway in a well known model organism – the addressed questions of consistent embedding of models with changed topologies can become more complex and unintuitive for other species and less well studied pathways. To have an algorithmic solution to the calculation of consistent leak fluxes will be a benefit in those cases and contribute to the reproducibility and transparency of the model development process.

Integrative Biology Accepted Manuscript

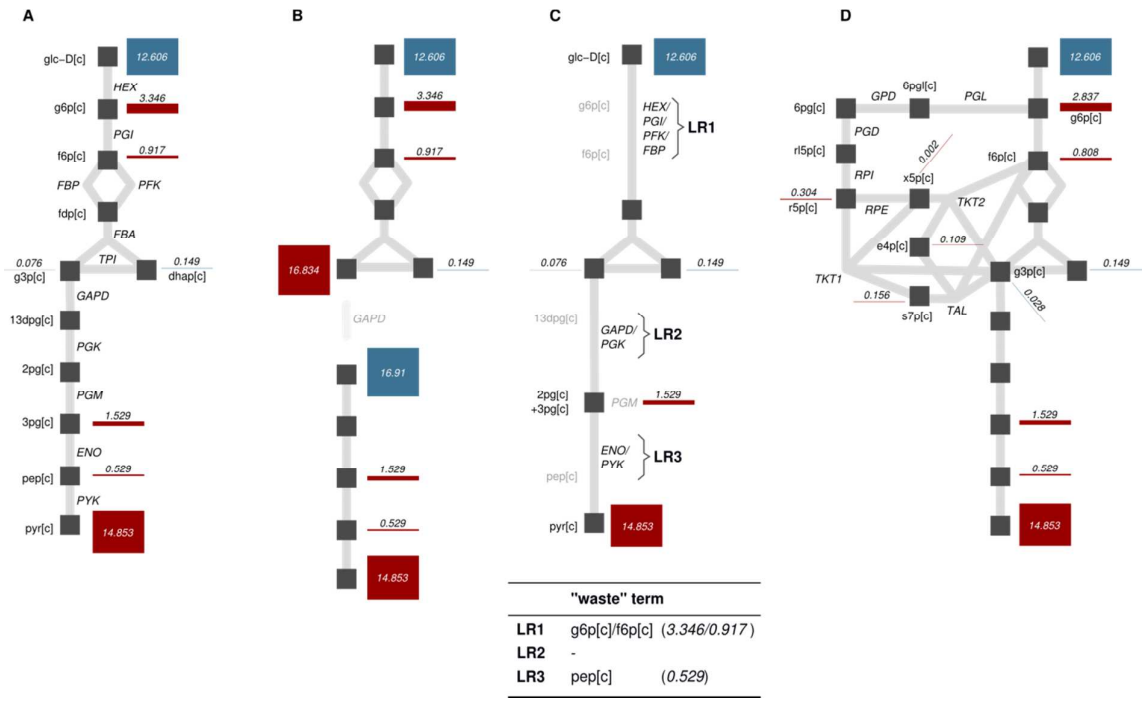


Figure 6: Calculating biomass leaks for split (**B**), simplified (**C**) and extended (**D**) models of yeast glycolysis (A). Bars show the magnitude of the leak flux of the corresponding dynamic metabolite (red – consuming, negative flux; blue – producing, positive flux)

Integrative Biology Accepted Manuscript

5 Discussion

We present an intuitive workflow to transfer the broad knowledge collected in genome scale metabolic models to smaller network parts of special interest, which allows investigating them in a cellular context. The workflow reproducibly generates a biomass drain reaction for a given dynamic metabolic model as well as biologically exact stoichiometries for lumped reactions in a specific biological scenario. The presented method has the potential to improve dynamic metabolic modeling for various applications since DDMs in matched cellular context will have higher predictive power.

The presented method helps to improve existing DDM that hitherto largely lacked contextualizing information. We have shown that the cellular context, environmental conditions as well as species specificity have major influence on the demand from DDMs. Metabolic models, in contrast to isolated signaling models, have to fulfill mass balance constraints in addition to show dynamics matching experimental data – an additional requirement impeding model calibration. Biomass backtracking is a first step into resolving this problem by fixing biologically meaningful boundary fluxes for the regions of the metabolic network covered by a DDM – thereby reducing the feasible parameter space and improving biological accuracy. The integration of various types of experimental data such as gene expression data ([45,46] and others), metabolomics and proteomics measurements[47], thermodynamic constraints[48,49], metabolic flux measurements and others[50] used to constrain the solution space of the GSM, can thereby directly be transmitted to the DDM to narrow the feasible parameter space.

A second focus of application is the iterative development of models, where improved model versions will still need to fulfill the overall cellular context. In the context of modular integrative or even whole cell modeling[1,3,6,51,52] it is essential that the submodels are matching, i.e. that all interface fluxes are consistent with each other and the overall cellular requirements. Biomass backtracking allows to calculate leak fluxes for several modular DDMs based on the same GSM and is hence able to assure a consistent cellular context. The reproducibility and ease of adaption to new model structures and data availabilities is a mayor benefit in large modular modeling projects.

In contrast to other methods[18–21] linking DDMs and GSMs, we only used the very strict steady state assumption to generate the stoichiometry for the biomass reaction of the dynamic model. Applicants can then define appropriate kinetics for this reaction and include it in the dynamic simulation. In other words, we generated a simple 'transfer function' of external perturbations or perturbations in the cell state to the pathway of interest, even if the impact point is far away and the effect unintuitive, but do not restrict the dynamic nature of the models more than necessary to account for the correct drain of biomass precursors. This contributes to reduction of the feasible parameter space for dynamic model fitting to a biologically feasible flux scenario.

As pointed out above, the effort required to map the DDM on the GSM strongly depends on the curation level and the annotation density of the two models. The results of the backtracking heavily depend on the quality of the used GSMs; the more exact they are, the more biologically relevant will the leak reaction be, especially if the GSM flux distribution is consistent with further experimental data and thermodynamic constraints. Another crucial point is the biomass composition itself: Many GSMs rely on a small set of measured data for the setup of the biomass reaction in their reconstruction. Often data are transferred from other organisms and hence lack the desired species specificity. Fixed stoichiometry biomass reactions do, moreover, not account for changing composition during nutrient shifts, cell cycle progression, stress or changes in growth rates. Biomass

Integrative Biology Accepted Manuscript

backtracking allows to generate distinct drain reactions for each component of the cellular biomass and to add specific regulation to them. This way, ratios of components can be easily shifted in accordance with biological context.
Our approach still represents a merely static mapping, i.e. the dynamics of the detailed models do not feedback on the composition of the biomass drain or the flux distribution of the GSM. A next step would be to develop new methods to more precisely couple dynamic simulations with constraint models of metabolism and simulate them together.

Backtracking paves the way for further application, e.g. in drug development. In complex diseases, systemic effects are crucial, but not intuitively covered by DDMs. Backtracking enables to integrate, e.g., cell type or even patient specific data into DDMs, increasing their predictive value in their specific context. For less explored systems of emerging biological interest, e.g. distinct areas of microbial metabolism such as lipid metabolism or sterol biosynthesis, biomass backtracking holds the potential to connect modeling hypotheses to the current state of knowledge on the surrounding metabolic system. This will ease model calibration as described above and increase the predictive power for example for the model driven development of novel drug targets[53,54].

6) Acknowledgements

We thank the EU SystemTB project for funding (EK), Jens Hahn for help with the Hynne model and Ali Saitov for support in testing the methods.

Integrative Biology Accepted Manuscript

7 References

1. Carrera, J. *et al.* An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of Escherichia coli. *Mol. Syst. Biol.* **10,** 735 (2014).

2. Wodke, J. a H. *et al.* Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. *Mol. Syst. Biol.* **9,** 653 (2013).

3. Petelenz-Kurdziel, E. *et al.* Quantitative Analysis of Glycerol Accumulation, Glycolysis and Growth under Hyper Osmotic Stress. *PLoS Comput. Biol.* **9,** e1003084 (2013).

4. Warburg, O. On the Origin of Cancer Cells. *Science (80-.).* **123,** 309–314 (1956).

5. Keasling, J. Manufacturing molecules through metabolic engineering. *Science (80-.).* **330,** 1355–8 (2010).

6. Klipp, E., Nordlander, B., Krüger, R., Gennemark, P. & Hohmann, S. Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.* **23,** 975–82 (2005).

7. Van Heerden, J. H. *et al.* Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science* **343,** 1245114 (2014).

8. Hynne, F., Danø, S. & Sørensen, P. G. Full-scale model of glycolysis in Saccharomyces cerevisiae. *Biophys. Chem.* **94,** 121–163 (2001).

9. Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K. & Reuss, M. Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnol. Bioeng.* **79,** 53–73 (2002).

10. Teusink, B. *et al.* Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.* **267,** 5313–29 (2000).

11. Achcar, F., Kerkhoven, E. J., Bakker, B. M., Barrett, M. P. & Breitling, R. Dynamic modelling under uncertainty: the case of Trypanosoma brucei energy metabolism. *PLoS Comput. Biol.* **8,** e1002352 (2012).

12. Singh, V. K. & Ghosh, I. Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in Mycobacterium tuberculosis, and its application to assessment of drug targets. *Theor. Biol. Med. Model.* **3,** 27 (2006).

13. Reed, M. C. *et al.* A mathematical model of glutathione metabolism. *Theor. Biol. Med. Model.* **5,** 8 (2008).

14. Messiha, H., Kent, E., Malys, N. & Carroll, K. Enzyme characterisation and kinetic modelling of the pentose phosphate pathway in yeast. *J Prepr.* (2014).

15. Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L. & Palsson, B. Ø. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* **7,** 129–43 (2009).

16. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28,** 245–8 (2010).

Integrative Biology Accepted Manuscript

17. Bordbar, A., Monk, J. M., King, Z. a & Palsson, B. O. Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* **15,** 107–20 (2014).

18. Mahadevan, R., Edwards, J. S. & Doyle, F. J. Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophys. J.* **83,** 1331–40 (2002).

19. Willemsen, a M. *et al.* MetDFBA: incorporating time-resolved metabolomics measurements into dynamic flux balance analysis. *Mol. Biosyst.* **11,** 137–145 (2014).

20. Chowdhury, A., Zomorrodi, A. R. & Maranas, C. D. k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.* **10,** e1003487 (2014).

21. Stanford, N. J. *et al.* Systematic construction of kinetic models from genome-scale metabolic networks. *PLoS One* **8,** (2013).

22. Jouhten, P., Wiebe, M. & Penttilä, M. Dynamic flux balance analysis of the metabolism of Saccharomyces cerevisiae during the shift from fully respirative or respirofermentative metabolic states to anaerobiosis. *FEBS J.* **279,** 3338–54 (2012).

23. Chelliah, V., Laibe, C. & Le Novère, N. BioModels Database: A Repository of Mathematical Models of Biological Processes. *Methods Mol. Biol.* **1021,** 189–99 (2013).

24. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19,** 524–531 (2003).

25. Becker, S. a *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* **2,** 727–38 (2007).

26. Keating, S. M., Bornstein, B. J., Finney, A. & Hucka, M. SBMLToolbox: an SBML toolbox for MATLAB users. *Bioinformatics* **22,** 1275–7 (2006).

27. Hastings, J. *et al.* The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.* **41,** D456–63 (2013).

28. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27,** 29–34 (1999).

29. Ibarra, R. U., Edwards, J. S. & Palsson, B. O. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420,** 186–9 (2002).

30. Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. & Sauer, U. Multidimensional optimality of microbial metabolism. *Science* **336,** 601–4 (2012).

31. Lewis, N. E. *et al.* Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6,** 390 (2010).

32. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5,** 264–276 (2003).

33. Aung, H. W., Henry, S. a & Walker, L. P. Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Ind. Biotechnol. (New Rochelle. N. Y).* **9,** 215–228 (2013).

Integrative Biology Accepted Manuscript

34. Schulze, U. Anaerobic Physiology of Saccharomyces Cerevisiae. (1995).

35. Van Eunen, K., Kiewiet, J. a L., Westerhoff, H. V. & Bakker, B. M. Testing biochemistry revisited: How in vivo metabolism can be understood from in vitro enzyme kinetics. *PLoS Comput. Biol.* **8,** (2012).

36. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol. Syst. Biol.* **7,** 535 (2011).

37. Hoops, S. *et al.* COPASI--a COmplex PAthway SImulator. *Bioinformatics* **22,** 3067–74 (2006).

38. Jouhten, P. *et al.* Oxygen dependence of metabolic fluxes and energy generation of Saccharomyces cerevisiae CEN.PK113-1A. *BMC Syst. Biol.* **2,** 60 (2008).

39. Archer, C. T. *et al.* The genome sequence of E. coli W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of E. coli. *BMC Genomics* **12,** 9 (2011).

40. Jamshidi, N. & Palsson, B. Ø. Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Syst. Biol.* **1,** 26 (2007).

41. Henry, C. S., Zinner, J. F., Cohoon, M. P. & Stevens, R. L. iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. *Genome Biol.* **10,** R69 (2009).

42. Anderson, J., Chang, Y. C. & Papachristodoulou, A. Model decomposition and reduction tools for large-scale networks in systems biology. *Automatica* **47,** 1165–1174 (2011).

43. Radulescu, O., Gorban, A. N., Zinovyev, A. & Noel, V. Reduction of dynamical biochemical reactions networks in computational biology. *Front. Genet.* **3,** 131 (2012).

44. Rao, S., van der Schaft, A., van Eunen, K., Bakker, B. M. & Jayawardhana, B. A model reduction method for biochemical reaction networks. *BMC Syst. Biol.* **8,** 52 (2014).

45. Colijn, C. *et al.* Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. *PLoS Comput. Biol.* **5,** e1000489 (2009).

46. Covert, M. W. & Palsson, B. Ø. Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *J. Biol. Chem.* **277,** 28058–64 (2002).

47. Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E. & Shlomi, T. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* **26,** i255–60 (2010).

48. Hoppe, A., Hoffmann, S. & Holzhütter, H.-G. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst. Biol.* **1,** 23 (2007).

49. Kümmel, A., Panke, S. & Heinemann, M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.* **2,** 2006.0034 (2006).

50. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10,** 291–305 (2012).

Integrative Biology Accepted Manuscript

51. Karr, J. R. *et al.* A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* **150,** 389–401 (2012).

52. Lee, J. M., Min Lee, J., Gianchandani, E. P., Eddy, J. a. & Papin, J. a. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.* **4,** e1000086 (2008).

53. Klipp, E., Wade, R. C. & Kummer, U. Biochemical network-based drug-target prediction. *Curr. Opin. Biotechnol.* **21,** 511–6 (2010).

54. Schulz, M., Bakker, B. M. & Klipp, E. TIde: a software for the systematic scanning of drug targets in kinetic network models. *BMC Bioinformatics* **10,** 344 (2009).

Integrative Biology Accepted Manuscript